Journal of Official Statistics, Vol. 24, No. 4, 2008, pp. 557–569

On Some Common Practices of Systematic Sampling

Li-Chun Zhang[1]

Systematic sampling is a widely used technique in survey sampling. It is easy to execute, whether the units are to be selected with equal probability or with probabilities proportional to auxiliary sizes. It can be very efficient if one manages to achieve favourable stratification effects through the listing of units. The main disadvantages are that there is no unbiased method for estimating the sampling variance, and that systematic sampling may be poor when the ordering of the population is based on inaccurate knowledge. In this article we examine an aspect of the systematic sampling that previously has not received much attention. It is shown that in a number of common situations, where the systematic sampling has on average the same efficiency as the corresponding random sampling alternatives under an assumed model for the population, the sampling variance fluctuates much more with the systematic sampling. The use of systematic sampling is associated with a risk that in general increases with the sampling fraction. This can be highly damaging for large samples from small populations in the case of single-stage sampling, or large subsamples from small subpopulations in the case of multi-stage sampling.

Key words: Statistical decision; second-order Bayes risk; robust design; panel survey.

1. Introduction

Systematic sampling has a long tradition in survey sampling (e.g., Madow and Madow 1944, Madow 1949, 1953). When applied to a list of units it is known as the "every *k*th rule", where *k* refers to the sampling interval. Where the ordering of the units is conceivably uncorrelated with the survey variable of interest, or contains at most a mild stratification effect, the systematic sampling is generally considered as a convenient substitute for simple random sampling "with little expectation of a gain in precision" (Cochran 1977, p. 229). The same holds for sampling within strata or subsampling under a multi-stage sampling design. By a modification (Madow 1949) where the sampling interval is calculated in terms of an accumulated auxiliary total, the systematic sampling can be used to select a πps sample with great ease.

In situations where auxiliary information is available for partial ordering of the population, it is more natural to compare systematic sampling with stratified random sampling. The systematic sampling is more convenient especially because it is not subjected to restrictions either on the number of auxiliary variables or on the number of levels each of them may take. So there is less need for variable selection than may be the

[1] Statistics Norway, Kongens gate 6, P.B. 8131-Dep, N-0033 Oslo, Norway. Email: lcz@ssb.no
Acknowledgements: I am very grateful to two referees for pointing out a technical mistake in a previous version of the article, and for suggesting a more elegant derivation of the second-order Bayes risks under the homogeneous population model.

case when it comes to stratified random sampling. By using many auxiliary variables the systematic sampling can introduce greater balance into the sample, although a more parsimonious stratified sampling design may well be just as efficient.

It is important to be clear that when speaking above of the efficiency of systematic sampling, we are referring to its sampling variance *in expectation*. Take for instance the case where systematic sampling is applied to a fixed, conceivably random ordering of a given population. The sampling variance, denoted by V_{sys}, is based on only k possible systematic samples, and is either larger or smaller than that of simple random sampling for the given population, denoted by V_{srs}. There are two results which show that V_{sys} may be equal to V_{srs} in expectation. In the first case, considering the fixed ordering to be randomly chosen from all $N!$ possible permutations of the N units of a finite population, Madow and Madow (1944) showed that $E(V_{sys}) = V_{srs}$, where the expectation is taken over all permutations. Notice that V_{srs} is a constant of the permutations. Example 3.4.2 of Särndal et al. (1992) provides a simple illustration of how greatly V_{sys} may vary for different population orderings. In the second case, we regard the ordering of the population as fixed, and the associated values of interest as realizations of independent random variables with constant mean (Cochran 1977, Theorem 8.5). It can then be shown that $E(V_{sys}) = E(V_{srs})$, with the expectation being over all possible finite populations under the assumed model.

To clarify the choice between systematic and simple random sampling in the situation above, we rephrase it as a decision problem. Let $\theta = (y_1, \ldots, y_N)^T$ be the vector of variables of interest, where $U = (1, \ldots, N)$ is a particular ordering of the units prepared for the systematic sampling. Given θ and U, we can choose to draw a systematic sample that depends on U, or we can choose to draw a simple random sample that does not depend on U. These are the two decision rules, or actions, available to us, denoted by $\delta = \text{SYS}$ and $\delta = \text{SRS}$, respectively. Let the sampling variance of the sample mean be our loss function, denoted by $L(\theta, \delta) = V_{\delta}(\bar{y}_s)$, where s denotes the selected sample and $\bar{y}_s$ the sample mean. Notice that this is a no-data problem, so that the frequentist risk of δ is equal to the loss function (Berger 1985), denoted by $R(\theta, \delta) \equiv L(\theta, \delta)$. Now, depending on the actual θ and U, $R(\theta, \text{SYS}) = V_{sys}$ may be greater or less than $R(\theta, \text{SRS}) = V_{srs}$, i.e., neither of them is R-better than the other for all θ. Indeed, $R(\theta,\delta)$ can be arbitrarily large as long as there is no limit on how much variation θ can have, such that e.g., the minimax principle is not applicable without further restrictions.

It is possible to invoke other decision principles. For instance, denote by $r(\pi, \delta) = E_{\theta}(V_{\delta}(\bar{y}_s))$ the Bayes risk of δ with respect to some assumed distribution of θ, denoted by $\pi(\theta)$, which is the expected sampling variance induced by δ in this case. Then, according to the Bayes risk principle, the decision rule SYS may be preferred to SRS if $r(\pi, \text{SYS}) < r(\pi, \text{SRS})$. As we have seen, $r(\pi, \text{SYS}) = r(\pi, \text{SRS})$ under the two models of θ above, so the choice between the two actions cannot be settled based on the Bayes risk principle alone.

Additional criteria of cost or easiness in execution can be used to motivate the choice of SYS in practice. Now, due to developments in computational power and alternative random sampling techniques, easiness in sample selection is no longer a valid argument in favour of systematic sampling. Using a computer one can draw a simple random sample as easily as a systematic sample. The same goes for πps sampling. For instance, sequential Poisson sampling (SPS, Ohlsson 1998) is easy to implement, yielding an approximate πps sample with a fixed sample size. Yet one needs to keep in mind the difference between ease of sample selection and ease of sample collection. For instance, in forestry surveys it is still easier to count every tenth tree than to consult a list of random numbers in the field.

Meanwhile, easiness in execution counts only if there are no other, more important decision principles that can be used to distinguish between the two actions. So we need to ask the following question: Is there any other reasonable decision principle that we may follow in this case, apart from e.g., the minimax principle and the Bayes risk principle?

The situation we are considering here has an analogy in Utility theory. Suppose that one is offered a 50–50 lottery between 0 and 100 pounds. The expected utility is 50 pounds. It is unlikely, however, that one is entirely indifferent between accepting the lottery and accepting 50 pounds for sure. One is risk averse if one prefers to accept the 50 pounds for sure than to enter the lottery; whereas one is risk prone if one prefers to enter the lottery instead (French 1986). For statistical decisions, however, we can motivate the same kind of distinction without reference to the lottery scenario. Let

$$d(\pi, \delta) = V_\theta(R(\theta, \delta)) \tag{1}$$

be the *second-order Bayes risk* of a decision rule δ with regard to $\pi(\theta)$. While the (first-order) Bayes risk is the expectation of the risk with regard to $\pi(\theta)$, the second-order Bayes risk is its variance. It is nonnegative by definition. In the case of zero second-order Bayes risk, the risk of a decision rule is the same regardless of the value of θ. The smaller the second-order Bayes risk, the more robust a decision rule is as θ varies. A decision rule δ is preferred to another δ' according to the *robust Bayes decision principle* if

$$r(\pi, \delta) = r(\pi, \delta') \quad \text{and} \quad d(\pi, \delta) < d(\pi, \delta') \tag{2}$$

That is, provided two rules have the same expected risk, we will choose the one that has less variation around the expected risk, on the ground of its robustness towards θ.

In the situation above, we have two sampling designs to choose from, which have the same Bayes risk under the assumed $\pi(\theta)$. The second-order Bayes risk is $d(\pi, \delta) = V_\theta(V_\delta(\bar{y}_s))$. It follows that if we choose between SYS and SRS according to the robust Bayes decision principle, we will have tighter control over the actual sampling variance over all possible θ. Notice that the second-order Bayes risk is a measure of robustness *given* $\pi(\theta)$. It is different from robustness towards misspecification of $\pi(\theta)$, which is a standard robustness concept in statistical decision theory. A decision rule δ may be preferred to another δ' according to the robust Bayes decision principle provided the conditions in (2) hold based on the assumed $\pi(\theta)$. Whereas what happens to the choice as $\pi(\theta)$ varies is another robustness concern, i.e., robustness towards misspecification of $\pi(\theta)$. A numerical illustration will be provided in Section 4 where both types of robustness are brought into consideration at the same time.

In the rest of the article we will be dealing with two issues. Firstly, we will show theoretically as well as by simulations that systematic sampling has greater second-order Bayes risk than the corresponding random sampling alternatives in all the situations mentioned at the beginning of this introduction. We believe that this is due to the fact that systematic sampling with a fixed list is cluster sampling, whereas clustering is removed under the random sampling alternatives. Our approach is based on population models, where we fix the ordering of the population and consider the values of interest as realized random variables under some assumed population model. This seems to be more in accordance with the practice of systematic sampling where the ordering is typically given once and for all. Notice that the problem of second-order Bayes risk is not the same as that of the instability of the variance estimator (Raj 1965), but a similar issue may arise because of the clustering due to systematic sampling. Secondly, we investigate possible consequences of ignoring the robust Bayes decision principle, i.e., choosing systematic sampling in spite of knowing that it has greater second-order Bayes risks. In particular, by simulations based on the Norwegian Census and Labour Force data, we show that the use of systematic sampling in panel surveys causes the estimates of changes in an autocorrelated population to vary considerably in precision over time, which is a fault that can not be overlooked in panel surveys. A summary will be given at the end.

2. Homogeneous Populations

Consider first equal-probability systematic sampling from a fixed population ordering that may be considered as uncorrelated with the variable of interest. Let the sample size be n, and let the sampling interval be k. For simplicity we assume that k is naturally an integer satisfying $N = nk$. Denote by s_m the mth systematic sample, i.e., $s_m = \{m, m+k, m+2k, \ldots, m+(n-1)k\}$. Let $\bar{y}_m$ be the corresponding sample mean, which is an unbiased estimator of the population mean, denoted by $\bar{Y} = \sum_{i \in U} y_i / N$. The sampling variance of $\bar{y}_m$ is given as $V_{sys} = k^{-1} \sum_{m=1}^{k} (\bar{y}_m - \bar{Y})^2$, which may or may not exceed the variance of the simple random sample mean, denoted by $V_{srs} = (n^{-1} - N^{-1})\sigma^2$, where $\sigma^2 = \sum_{i \in U} (y_i - \bar{Y})^2 / (N-1)$.

As mentioned before, there are two results which show that SRS and SYS have the same Bayes risk, i.e., $E_\theta(V_{srs}) = E_\theta(V_{sys})$. We shall investigate their second-order Bayes risks under the following homogeneous model for the population

$$E(y_i) = \mu \quad \text{and} \quad E((y_i - \mu)^r) = \mu_r < \infty \tag{3}$$

for $i \in U$, where for simplicity we write E instead of E_θ, and y_i is independent of y_j for $i \neq j$. We have $E(V_{srs}) = E(V_{sys}) = (1/n - 1/N)\mu_2 = (1 - 1/k)\mu_{2,n}$, where $\mu_{r,n}$ denotes the rth central moment of $\bar{y}_m$. This is a special case of the more general Theorem 8.5 of Cochran (1977), where the model variance of y_i is allowed to vary over the units.

Moreover, based on a result of the variance of the empirical variance (e.g., Wetherill 1981), we obtain

$$V(V_{srs}) = \left(\frac{1}{n} - \frac{1}{N}\right)^2 V(\sigma^2) = \left(1 - \frac{1}{k}\right)^2 \mu_{2,n}^2 \left(\frac{2}{N-1} + \frac{\gamma_2}{N}\right)$$

where $\gamma_2 = \mu_4/\mu_2^2 - 3$ is the coefficient of kurtosis of y_i. Similarly, we have

$$V(V_{sys}) = \left(\frac{k-1}{k}\right)^2 V\left(\frac{1}{k-1}\sum_{m=1}^{k}(\bar{y}_m - \bar{Y})^2\right) = \left(1 - \frac{1}{k}\right)^2 \mu_{2,n}^2 \left(\frac{2}{k-1} + \frac{\gamma_{2,n}}{k}\right)$$

where $\gamma_{2,n} = \gamma_2/n$ is the coefficient of kurtosis of $\bar{y}_m$. It follows that the coefficients of variation (CV) of V_{srs} and V_{sys} are, respectively,

$$CV(V_{srs}) = \left(1 - \frac{1}{k}\right)\left(\frac{2}{N-1} + \frac{\gamma_2}{N}\right)^{\frac{1}{2}} \quad \text{and}$$
$$CV(V_{sys}) = \left(1 - \frac{1}{k}\right)\left(\frac{2}{k-1} + \frac{\gamma_2}{N}\right)^{\frac{1}{2}} \tag{4}$$

Provided the population is large enough we have $CV(V_{sys}) \approx (1 - 1/k)\sqrt{2/(k-1)}$, which increases with the sampling fraction. For example, the overall sampling fraction is about 1/140 in the Norwegian Labour Force Survey (LFS), such that $CV(V_{sys}) \approx 12\%$. In comparison, $CV(V_{srs}) = O(1/\sqrt{N})$ and the second-order Bayes risk of simple random sampling is negligible. Drawing systematic samples from a seemingly random, but fixed list of population is a haphazard business without expectation of gains compared to simple random sampling. One simply has less control over the actual sampling variance, which may considerably deviate from its expectation. The same holds for stratified systematic sampling compared to stratified simple random sampling. In two-stage sampling where systematic sampling is used for subsampling of units within a primary sampling unit (PSU), what counts for the second-order Bayes risk is the within-PSU sampling fractions. Thus, systematic sampling can have a large second-order Bayes risk in the case of multi-stage sampling, even though the overall sampling fraction may be low.

3. Ratio Regression Populations

Consider now the situation for systematic πps sampling. In this case the "every *k*th" rule is applied to the cumulated total of an auxiliary variable, denoted by x_i for $i \in U$. Any fixed list U can be used. For simplicity we assume that x_i is an integer. Let $X = \sum_{i \in U} x_i$. The interval length is then given by $k = X/n$, where again we assume that k is naturally an integer. Looked at the other way around, equal-probability systematic sampling becomes systematic πps sampling with $x_i \equiv 1$. The unit i may appear in x_i different systematic samples. We assume that the inclusion probability is such that $\pi_i = nx_i/X < 1$ for all $i \in U$. Based on any systematic πps sample, denoted by s_m for $m = 1, \ldots, k$, the estimator of Y is

$$\hat{Y}_m = \sum_{i \in s_m} \frac{y_i}{\pi_i} = \frac{X}{n} \sum_{s_m} b_i = X\bar{b}_m \quad \text{for} \quad b_i = \frac{y_i}{x_i} \quad \text{and} \quad \bar{b}_m = \sum_{i \in s_m} \frac{b_i}{n}$$

.

We have $E_{sys}(\bar{b}_m) = Y/X$, and $E_{sys}(\bar{y}_m) = Y$, and

$$V_{sys}(\hat{Y}_m) = X^2 V_{sys}(\bar{b}_m) = X^2 \left\{ \frac{1}{k} \sum_{m=1}^{k} (\bar{b}_m - Y/X)^2 \right\}$$

Now, $\bar{b}_m$ is the best linear unbiased estimator (BLUE) of β under the following model

$$y_i = x_i\beta + x_i\varepsilon_i \tag{5}$$

where $E(\varepsilon_i) = 0$, and $V(\varepsilon_i) = \mu_2$, and $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j \in U$, i.e., a ratio regression model with residual variance proportional to x_i^2. We have

$$V_{sys}(\bar{b}_m) = \frac{1}{k}\sum_{m=1}^{k}(\bar{b}_m - \beta)^2 - (Y/X - \beta)^2 \doteq \frac{1}{k}\sum_{m=1}^{k}(\bar{b}_m - \beta)^2 = \frac{1}{k}\sum_{m=1}^{k}\bar{\varepsilon}_m^2 \overset{def}{=} Z$$

where $\bar{\varepsilon}_m = \sum_{i\in s_m}\varepsilon_i/n$, provided the population is large enough so that the term $(Y/X - \beta)^2$ is of a lower order compared to Z. We have $E(Z) = \mu_{2,n}$. Moreover,

$$Z^2 = \frac{1}{k^2}\left(\sum_{m=1}^{k}\bar{\varepsilon}_m^4 + \sum_{m,p;p\neq m}\bar{\varepsilon}_m^2\bar{\varepsilon}_p^2\right)$$

where $\bar{\varepsilon}_m$ and $\bar{\varepsilon}_p$ are not necessarily independent of each other because some units may appear both in s_m and s_p. However,

$$E\left(\bar{\varepsilon}_m^2\bar{\varepsilon}_p^2\right) = n^{-4}\sum_{\substack{(i_1,i_2)\in s_m,\\ (j_1,j_2)\in s_p}} E(\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{j_1}\varepsilon_{j_2})$$

where $E(\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{j_1}\varepsilon_{j_2})$ is not zero, indeed positive only if it is of the form $E\left(\varepsilon_i^4\right)$ or $E\left(\varepsilon_i^2\varepsilon_j^2\right)$. Let s_{mp} denote the intersection of s_m and s_p. We have

$$\sum_{\substack{(i_1,i_2)\in s_m,\\ (j_1,j_2)\in s_p}} E(\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{j_1}\varepsilon_{j_2}) = E\left(2\sum_{i\neq j\in s_{mp}}\varepsilon_i^2\varepsilon_j^2 + \sum_{\substack{i\in s_m,\\ j\in s_p}}\varepsilon_i^2\varepsilon_j^2\right)$$

The first term arises by taking $(i_1, i_2) = (j_1, j_2)$ and $(i_1, i_2) = (j_2, j_1)$ with $i_1 \neq i_2$, and the second term by taking $i_1 = i_2$ and $j_1 = j_2$. Let s_m^p denote the units of s_m that are not included in s_p, and let s_p^m denote the units of s_p that are not included in s_m. We have

$$\sum_{i\in s_m,\, j\in s_p}\varepsilon_i^2\varepsilon_j^2 = \sum_{i,j\in s_{mp}}\varepsilon_i^2\varepsilon_j^2 + \sum_{i\in s_{mp},\, j\in s_p^m}\varepsilon_i^2\varepsilon_j^2 + \sum_{i\in s_{mp},\, j\in s_m^p}\varepsilon_i^2\varepsilon_j^2 + \sum_{i\in s_p^m,\, j\in s_m^p}\varepsilon_i^2\varepsilon_j^2$$

and

$$\sum_{i,j\in s_{mp}}\varepsilon_i^2\varepsilon_j^2 = \sum_{i\in s_{mp}}\varepsilon_i^4 + \sum_{i\neq j\in s_{mp}}\varepsilon_i^2\varepsilon_j^2$$

Thus, we obtain

$$n^4 E\left(\bar{\varepsilon}_m^2 \bar{\varepsilon}_p^2\right) = 2\sum_{i\neq j\in s_{mp}} \mu_2^2 + \left(\sum_{i\in s_{mp}} \mu_4 - \sum_{i\in s_{mp}} \mu_2^2\right)$$

$$+ \left(\sum_{i\in s_{mp}} \mu_2^2 + \sum_{i\neq j\in s_{mp}} \mu_2^2 + \sum_{\substack{i\in s_{mp},\\ j\in s_p^m}} \mu_2^2 + \sum_{\substack{i\in s_{mp},\\ j\in s_m^p}} \mu_2^2 + \sum_{\substack{i\in s_p^m,\\ j\in s_m^p}} \mu_2^2\right)$$

$$= 2c_{mp}(c_{mp}-1)\mu_2^2 + c_{mp}(\gamma_2+2)\mu_2^2 + n^2\mu_2^2 = \mu_2^2\left(n^2 + c_{mp}\gamma_2 + 2c_{mp}^2\right)$$

where c_{mp} is the number of common units in s_m and s_p, and the two terms that involve c_{mp} exist only if s_{mp} is not empty. Denote by $\mu_{4,n}$ the fourth central moment of $\bar{\varepsilon}_m$. We now have

$$V(Z) = \frac{1}{k}\left(\mu_{4,n} - \mu_{2,n}^2\right) + \frac{\mu_{2,n}^2}{k^2n^2}\left(\gamma_2\sum_{m\neq p} c_{mp} + 2\sum_{m\neq p} c_{mp}^2\right) = \mu_{2,n}^2 \lambda_n(\gamma_2) \tag{6}$$

where

$$\lambda_n(\gamma_2) = 2\left(\frac{1}{k} + \frac{\sum_{m\neq p} c_{mp}^2}{X^2}\right) + \gamma_2 \frac{\sum_{i=1}^N x_i^2}{X^2}$$

since $\mu_{4,n} = \mu_{2,n}^2(\gamma_{2,n}+3)$, and $X = nk$, and $\sum_{m\neq p} c_{mp} = \sum_{i=1}^N x_i(x_i-1)$. The term $\sum_{m\neq p} c_{mp}^2$ seems intractable in general, but it can be calculated given the ordering of the units. It follows that $CV(V_{sys}(\hat{Y}_m)) \doteq CV(Z) = \sqrt{\lambda_n(\gamma_2)}$.

Meanwhile, there are a variety of alternative random πps sampling methods. It is easily shown that in the case of Poisson sampling (PS), the CV of $V_{ps}(\hat{Y})$ is of the order $O(1/\sqrt{N})$ under Model (5). The result holds generally for any fixed-sized πps sampling design, provided its sampling variance is related to that of the PS through a finite population correction term.

4. A Numerical Illustration

For a numerical illustration of Results (4) and (6), let us consider sampling of 10 units from a population of 100, denoted by $U = \{1, 2, \ldots, 100\}$. The auxiliary variables are simply given as $x_i = i$. The survey variables y_i are to be simulated under the model below.

$$y_i = x_i + x_i^a \varepsilon_i \quad \text{where} \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \quad \text{and} \quad 0 \leq a \leq 1 \tag{7}$$

The conditional variance of y_i given x_i is thus equal to $x_i^{2a}\sigma^2$. In the case of $a = 0$, $y_i - x_i$ follows the homogeneous model (3). In the case of $a = 1$, we have Model (5) with $\beta = 1$, which can be used to motivate the πps sampling.

Consider first the πps sampling. Let $a = 0, 0.25, 0.5, 1$ and $\sigma = 0.01, 0.1$. Notice that σ cannot be too large here before negative y-values can be generated with nonnegligible probabilities, in which case the rationale for πps sampling would be doubtful. For each pair of (a,σ) we generate a population $\theta = (y_1, \ldots, y_{100})^T$, for which three sampling variances are calculated. The first one is for the systematic πps sampling. The second one is for the SPS, which is an approximate random πps sampling method. This is calculated by simple Monte Carlo. Finally, we calculate the asymptotic theoretical sampling variance of systematic πps sampling, with random permutation of θ before drawing a systematic sample, i.e.,

$$V_{asy} = \sum_{i=1}^{100} \pi_i \left(1 - \frac{n-1}{n}\pi_i\right)\left(\frac{y_i}{\pi_i} - \frac{Y}{n}\right)^2$$

(Hartley and Rao 1962), which can be used to benchmark the efficiency of the other two.

The simulations are repeated for 1,000 independently generated θ. The results are given in Table 1, where the relative efficiency (RE) refers to the ratio $E(V_\delta)/E(V_{asy})$. We notice the following. (I) It is seen that both the systematic πps sampling and the SPS achieve RE around 100%, such that the two sampling methods are equivalent with regard to the Bayes risk principle. (II) Under Model (5), i.e., $a = 1$, the CV of systematic πps sampling can be derived from (6). We have $\gamma_2 = 0$ given normality of ε_i and $\sum_{m \neq p} c_{mp}^2 = 2079500$ given U above, such that $\sqrt{\lambda_n(\gamma_2)} = 0.410$ which does not depend on σ^2. The theoretical CV is confirmed by the simulations. It is seen that systematic πps sampling has much greater second-order Bayes risk than random πps sampling. (III) The second-order Bayes risk varies little over σ given a. For $0.25 < a < 1$ the second-order Bayes risk of random πps sampling is almost a constant, and is considerably lower than that of the systematic πps sampling. The second-order Bayes risk of random πps sampling increases quickly as a gets close to 0, but remains lower than that of systematic πps sampling. In summary, random πps sampling is preferred according to the robust Bayes decision principle under Model (5), and the choice is robust towards departures from the assumption of $a = 0$, which is a form of misspecification of $\pi(\theta)$.

Consider next equal-probability systematic sampling. There is a general result which states that systematic sampling is more efficient than SRS provided that the within-sample

Table 1. Simulation results for πps sampling in percentage

Design		Relative Efficiency		CV of Sampling Variance		
		Systematic	SPS	Systematic	SPS	Theoretical
$a = 1$	$\sigma = 0.01$	98	100	41	16	16
	$\sigma = 0.1$	101	100	41	16	16
$a = 0.5$	$\sigma = 0.01$	101	100	34	15	15
	$\sigma = 0.1$	99	99	34	15	15
$a = 0.25$	$\sigma = 0.01$	100	99	28	18	18
	$\sigma = 0.1$	99	99	29	18	18
$a = 0$	$\sigma = 0.01$	100	99	37	36	35
	$\sigma = 0.1$	101	99	40	38	37

Relative efficiency: Ratio between average sampling variance of systematic πps (or SPS sampling) and average theoretical variance V_{asy}. CV of sampling Variance: Coefficient of variation of respective sampling variances.

variance is larger than the population variance, due to the following decomposition:

$$\sum_{i\in U}(y_i - \bar{Y})^2 = \sum_{m=1}^{k}\sum_{i\in s_m}(y_i - \bar{y}_m)^2 + \sum_{m=1}^{k} n(\bar{y}_m - \bar{Y})^2,$$

i.e., the variation within the k systematic samples and the variation between them. Since V_{sys} is proportional to the second component, it is minimized for a given θ when the first component is maximized. Based on the corresponding ordering of units, systematic sampling could potentially lead to gains in efficiency over simple random sampling. For instance, suppose the extreme case under Model (7) with $\sigma = 0$, i.e., $y_i = x_i$. The optimal ordering for a systematic sample of 10 units is to alternate between increasing and decreasing order every 10 units in the population (Särndal et al. 1992, Example 3.4.2), denoted by $U_{opt} = (1, \ldots, 10, 20, \ldots, 11, 21, \ldots, 30, 40, \ldots, 31, \ldots, 100, \ldots, 91)$, in which case $V_{sys}(\bar{y}_s) = 0$.

In practice, of course, one never knows y_i exactly. However, the ordering U_{opt} remains optimal under Model (7) with $a = 0$, now that $\bar{x}_m = X/N$ is a constant of sampling. The estimator based on an equal-probability systematic sample drawn from U_{opt} is given by

$$\hat{Y}_m = X + N\bar{\varepsilon}_m \quad \text{where} \quad X = \sum_{i=1}^{N} x_i \quad \text{and} \quad \bar{\varepsilon}_m = \sum_{i\in s_m}(y_i - x_i)/n$$

which is the same as the difference estimator (Särndal et al. 1992, Chapter 6.3) based on a simple random sample. In other words, the efficiency of systematic sampling based on U_{opt} can as well be achieved by the combined use of simple random sampling and difference estimator. Of course, the second-order Bayes risk of the latter strategy is only of order $O(1/\sqrt{N})$. The situation is illustrated in Table 2, where RE refers to $E(V_{sys})/E(V_{srs})$, and CV_δ is the CV of the actual sampling variance induced by δ = SYS and SRS, respectively. We notice the following. (I) As expected, the combined use of simple random sampling and difference estimator is as efficient as the optimal systematic sampling under Model (7) with $a = 0$. The systematic sampling becomes slightly less efficient under departures from the assumption $a = 0$, i.e., as a moves from 0 towards 1. (II) The second-order Bayes risk of systematic sampling is much greater than that of simple random sampling under the assumption $a = 0$. There is little variation in either for $0 \leq a \leq 1$. In summary, the combined use of simple random sampling and difference estimator is preferred to the optimal systematic sampling according to the robust Bayes

Table 2. Simulation results for equal-probability sampling in percentage: Systematic sampling based on U_{opt} vs. combined use of simple random sampling and difference estimator

	$a = 0$			$a = 0.5$			$a = 1$		
	RE	CV_{sys}	CV_{srs}	RE	CV_{sys}	CV_{srs}	RE	CV_{sys}	CV_{srs}
$\sigma = 0.01$	104	47	16	107	48	18	112	48	19
$\sigma = 0.1$	99	49	16	107	47	17	112	47	20

RE: Ratio between average systematic sampling variance and that of simple random sampling. CV: Coefficient of variation of respective sampling variances.

decision principle under Model (7) with $a = 0$, and the choice is robust towards departures from this assumption.

5. Systematic Sampling for Several Occasions

A systematic sample, once drawn, may be used for several subsequent occasions. Such a sample may constitute a group in a rotating panel design, such as that of the LFS in most countries. It can also form the core of a panel survey, with supplementary units added to the sample from time to time, in order to account for natural regeneration of the population. To simplify the discussion we assume here that a single systematic sample is drawn on the first occasion and used for all the subsequent occasions before it is abandoned, and that the population U remains the same throughout the period.

Results (4) and (6) apply then directly to the entire active period of the panel. More explicitly, let $\mathbf{y}_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})^T$ be the variables of interest associated with $i \in U$. Results (4) and (6) apply directly to any function of $\mathbf{y}_i$. For instance, suppose that $\mathbf{y}_i$ consists of 4 employment status measures in each of the 4 quarters of a calendar year from the LFS. The average yearly employment is given by the average of y_{i1} to y_{i4}. By drawing a systematic sample on the first occasion, one risks a variance fluctuation in the estimator of the average yearly employment rate as well as in any single quarter. Moreover, an important use of panel data is to estimate changes in the population. Let $z_i = (z_{i2}, \ldots, z_{iT})^T$, where $z_{it} = y_{it} - y_{i,t-1}$ for $t = 2, \ldots, T$, are the period-to-period changes. Again, Results (4) and (6) apply directly to any z_{it}, such that the estimation of change may have a high second-order Bayes risk due to systematic sampling.

The considerations above do not take into account possible strong autocorrelation among $\mathbf{y}_i$, which one often finds in natural populations. A conditional examination is needed in addition. Consider the simplest setting where $T = 2$, and y_{it} is a categorical variable such as the employment status. As a simple model of the dependence between y_{i1} and y_{i2} we assume Markov transition probability p_{ab} for $y_{i2} = b$ given $y_{i1} = a$, independent for $i \neq j \in U$. This amounts to a homogeneous population model (3) conditional on $y_{1i} = a$. The systematic sample mean of $z_i = y_{i2} - y_{i1}$ is given as

$$\bar{z}_m = \sum_{a;y_{i1}=a} \frac{n_a}{n} \bar{z}_{m,a}$$

where n_a is the number of units with $y_{i1} = a$ and $\bar{z}_{m,a}$ is the mean of change among them. Closed expression of the conditional variance $V_{sys}(\bar{z}_m | \{y_{i1}; i \in U\})$ appears intractable in general.

Instead, consider any ordering where the units are segmented according to the value of y_{i1}. Assume that N_a/k is naturally an integer for all a, where N_a is the number of units with $y_{i1} = a$ in the population. Both n_a and $\bar{y}_{m,t=1}$ then become constants of sampling, such that the variance of $\bar{z}_m$ is simply the variance of $\bar{y}_{m,t=2}$. Result (4) can now be applied to $\bar{y}_{m,a,t=2}$, i.e., within each segment of y_{i1} under the Markov transition model, such that the second-order Bayes risks of $\bar{y}_{m,a,t=2}$ given $\{y_{i1}; i \in U\}$ carries straight over to $\bar{z}_m$. Consideration of this special case suggests that the second-order Bayes risk of systematic sampling can be high for estimators of change in auto-correlated populations, also when the variance is evaluated conditionally. We shall examine this issue in a simulation study below.

6. Simulation: Labour Market Dynamics

We simulate the labour market dynamics using data from the Norwegian Census 2001 and the Norwegian LFS as follows. From the Census 2001, we obtain the employment status, classified as "Employed", "Unemployed" or "Not in the labour force", which is to be treated as the variable of interest in the population at $t = 1$. Next, from the LFS of the last quarter in 2004 and the first quarter in 2005, we observe a 3×3-transition matrix for the employment status between the two quarters. Using these Markov transition probabilities, we are able to simulate an employment status in the population at $t = 2$. The population within each of the 19 counties in Norway is sorted by municipality, age, sex, and the personal identification number (PIN), where the PIN may be considered as uncorrelated with the employment status of interest.

We consider four different strategies: (1) equal-probability systematic sampling at $t = 1$ and estimation based on direct weighting, denoted by Sys-Dir, (2) simple random sampling at $t = 1$ and estimation based on direct weighting, denoted by SRS, (3) proportionally allocated stratified random sampling with regard to sex and age (altogether 22 groups) followed by stratified estimation, denoted by Str-SRS, and (4) equal-probability systematic sampling followed by post-stratified estimation, with the 22 age-sex groups as post-strata, denoted by Sys-Pst.

The simulations are carried out separately for each of the 19 counties of Norway, reflecting the stratified design of the Norwegian LFS. A sample selected at $t = 1$ is also used at $t = 2$, and the within-county sample sizes are taken from the Norwegian LFS. The results are very similar for all the counties. Here we show only the situation for Østfold in Figure 1. Systematic sampling can in this case be considered as implicit stratification with

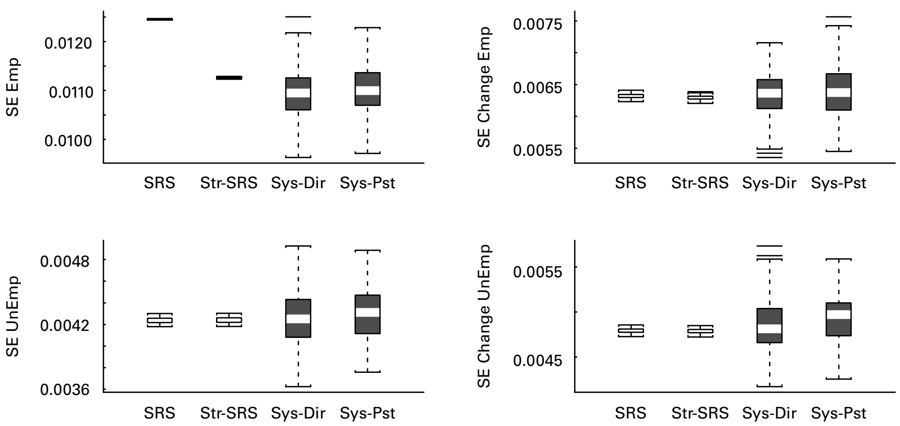

Fig. 1. Boxplot of standard error of (SE) of employment rate at t = 2 (Emp), change in employment rate (Change Emp), unemployment rate at t = 2 (UnEmp) and change in unemployment rate (Change Unemp) for county Østfold: direct weighting following simple random sampling (SRS), stratified random sampling with proportional allocation (Str-SRS), systematic sampling (Sys-Dir), and poststratified estimation following systematic sampling (Sys-Pst)

regard to municipality, age and sex. The stratification effects are notable only for employment rate at $t = 2$ (Emp), giving about 20% variance reduction compared to SRS. Most of the effect, however, can be achieved through stratification with regard to sex and age alone. Notice that stratification with regard to municipality in addition is unpractical due to the large number of strata. In all the other cases, no gains of efficiency can be expected from using systematic sampling.

It is seen that, while the second-order Bayes risks of SRS and Str-SRS are negligible for a population of this size (about 179 thousand persons), they are appreciable under systematic sampling also when the variances are evaluated conditionally as is done here. The CV of V_{sys} is 11.0% for Emp, 15.8% for Change Emp, 16.4% for UnEmp, and 15.4% for Change UnEmp. These are comparable to the theoretical unconditional value given by (4), which is approximately $\sqrt{2f} = \sqrt{2/134.6} = 12.2\%$ for Østfold. On certain occasions, therefore, the variance fluctuation may cancel out the expected stratification effect on the estimation of Emp. Notice that the second-order Bayes risk of systematic sampling cannot be reduced by means of post-stratification.

In particular, for the estimation of changes which is our primary concern here, the CV of systematic sampling variance is about the same as in the case of level estimation. Thus, the use of systematic sampling may cause the actual sampling variance of a change estimator to vary greatly over time. For instance, if the actual variance is 15% above the expectation between the first and second quarters, and it is 15% below the expectation between the second and third quarters, then the two change estimates have a difference of 30% in their sampling variances, caused by the use of systematic sampling alone. Now that the CV for the variance of either change estimator is about 15% here, this is hardly an unusual scenario. In the more extreme case of two standard deviations up or down from the expected sampling variance, the actual variance of one change estimator is almost twice (i.e., 1.3/0.7) that of the other. It is certainly undesirable to keep this as a feature of the sampling design.

7. Summary

In the above we introduced the concept of second-order Bayes risk and the robust Bayes decision principle. We have considered a number of situations where systematic sampling is commonly used as a substitute for alternative random sampling methods that are equally efficient. It is shown that the practice can induce large second-order Bayes risks, i.e., fluctuations in the actual sampling variance, both in cross-sectional and longitudinal survey sampling. This can be highly damaging for large samples taken from small populations, or large subsamples from small subpopulations. The use of systematic sampling for convenience is in such situations a haphazard business without any expectation of gains in efficiency.

Systematic sampling is also frequently applied outside the situations that we have considered. Cochran (1977) cited several examples, including the common use of one- or two-dimensional systematic sampling in forestry and land surveys. Such situations can be studied similarly to how it has been done here, but will require rather special population models containing correlations over both time and space, which are beyond the scope of this article. Moreover, as pointed out earlier, a systematic sample may still be easier to collect in such situations. Finally, the balance in a systematic sample may figure prominently in land or ecology surveys. For instance, an even spread of the sample sites may be considered more important than randomness of the sample for good spatial smoothing of the data.

We have studied systematic sampling from a statistical decision point of view, where the loss function is defined as the sampling variance of the survey estimator. Notice that, for model-based inference where the variance of an estimator is evaluated under the population model alone, the second-order Bayes risk of systematic sampling does not differ from that of an alternative random sampling method, provided the sampling is noninformative in both cases. Indeed, systematic sampling is sometimes considered to be useful as a first step in constructing various balanced samples (Valliant et al. 2000, Chapter 3).

8. References

Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, (2nd ed.). Springer. NY: New York.

Cochran, W.G. (1977). Sampling Techniques, (3rd ed.). New York: John Wiley and Sons.

French, S. (1986). Decision Theory: An Introduction to the Mathematics of Rationality. Chichester: Ellis Horwood Ltd.

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with Unequal Probabilities and Without Replacement. Annals of Mathematical Statistics, 33, 350–374.

Madow, W.G. (1949). On the Theory of Systematic Sampling, II. Annals of Mathematical Statistics, 20, 333–354.

Madow, W.G. (1953). On the Theory of Systematic Sampling, III. Annals of Mathematical Statistics, 24, 101–106.

Madow, W.G. and Madow, L.H. (1944). On the Theory of Systematic Sampling. Annals of Mathematical Statistics, 15, 1–24.

Ohlsson, E. (1998). Sequential Poisson Sampling. Journal of Official Statistics, 14, 149–162.

Raj, D. (1965). Variance Estimation in Randomized Systematic Sampling with Probability Proportionate to Size. Journal of the American Statistical Association, 60, 278–284.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). Finite Population Sampling and Inference. New York: Wiley.

Wetherill, B.G. (1981). Intermediate Statistical Methods: Chapman and Hall: London.

Received August 2006
Revised April 2008